CCD图像中宇宙线μ子甄选技术*

冯海霞[1,2,3]，陈建军[1,3*]，邓建榕[1]，赵永恒[1,2,3]

（1.中国科学院国家天文台，北京 100101；2.中国科学院大学，北京 100049；

3.中国科学院光学天文重点实验室（国家天文台），北京 100101）

摘 要：CCD（Charge Coupled Device）图像中的宇宙线是来自于外太空的高能粒子，它穿过大气层，与大气层的粒子发生相互作用形成次级粒子，最后落在CCD上。μ子是大气层宇宙线的主要成分，为了研究这些μ子的性质和变化规律，首先需要对图像中的μ子进行甄选。本文介绍一种快速有效地从CCD图像中甄选μ子的方法。采用拉普拉斯边缘检测法从CCD图像中提取宇宙线候选像素列表，去除其中的坏像素和噪声，然后使用凝聚层次聚类算法将宇宙线像素聚类成宇宙线事件，对宇宙线事件进行特征提取和分类，从中甄选宇宙线μ子。最后对甄选结果进行总结分析。

关键词：CCD；μ子；拉普拉斯边缘检测法；凝聚层次聚类；特征提取

中图分类号：TP391.41　　**文献标识码：**A　　**文章编号：**

宇宙线是来自于外太空的高能粒子。1912年，由奥地利地理学家赫斯在电离室对电流的测量中第一次发现。宇宙线中质子和氦核占98%，重核1%，电子1%，其他粒子微量[1]。当初级宇宙线进入地球大气层时，质子和氦核等高能粒子与大气中的原子核相互作用产生级联的广延大气簇射，生成大量的次级粒子。在使用CCD观测目标源时，这些粒子在CCD上留下不同的径迹，其中有电子、α粒子、X射线以及μ子等粒子。靠近大气底部的宇宙线主要是簇射产生的次级μ子。在CCD图像的宇宙线中，占据主要成分的是μ子。μ子和电子等带电粒子，在CCD上留下明显的特征径迹。电子大部分是由于环境中放射性同位素所产生的，并非来源于宇宙线（在参考文献[2]有详细介绍）。α粒子和X射线等其他粒子在CCD图像中图像特征呈点状，不易与CCD芯片本身产生的电子学噪声分开。因此研究CCD图像中的宇宙线，需要对宇宙线中的μ子进行甄选。

现如今，探测宇宙线的装置有很多，例如西藏羊八井国际宇宙线观测站、印度乌蒂的GRAPES-3、阿根廷的Pierre Auger Observatory以及四川的LHAASO（高海拔宇宙线观测站）等。这些装置采用一系列探测器阵列和望远镜来探测宇宙线粒子。本文利用LAMOST的CCD图像来探测μ子。LAMOST(The Large Sky Area Multi-Object Fiber Spectroscopic Telescope) 是中国目前最大口径的望远镜。参考文献[3]系统性地介绍了LAMOST的结构、原理及应用。

---

* **基金项目：**国家自然科学基金联合基金项目(U1631131)资助.

作者简介：冯海霞，女，硕士，研究方向：天文CCD图像数据处理。 Email: hxfeng@nao.cas.cn

通讯作者简介：陈建军，男，高级工程师，研究方向：天文技术方法、光谱数据处理。Email: jjchen@nao.cas.cn

LAMOST配置有16个光谱仪和32台4k×4k的CCD相机。CCD芯片大小为49.2mm×49.2mm，像素大小为12μm×12μm，工作温度在-100℃左右。从2011年先导巡天至今，已经积累了一百万多张CCD原始图像，每幅图像都有很多宇宙线事件，可以用来研究宇宙线的性质和分布。CCD图像中的宇宙线个数与图像的曝光时间有关，将曝光时间归一化到分钟。在单位时间内，CCD图像中的宇宙线事件个数范围为5～60。

CCD的工作原理是利用光电效应将光子转化为电子，电容器吸收释放的电子，然后将收集的电子转化为电压，最后以数字信号的形式输出。μ子有带负电和带正电两种，电荷数为1。它的质量约为106MeV，比电子重207倍。当它穿过CCD时，可以电离径迹上的硅原子，受到库伦散射和轫致辐射的影响比电子小，在CCD上留下可识别的直线径迹。

参考文献[2]介绍CCD图像中的宇宙线事件（指在CCD图像中一组有计数的连续像素）关于质心呈现不同的分布，结合宇宙线事件的形态特征甄选μ子。本文基于LAMOST数据，对CCD图像进行分析处理。根据μ子在CCD图像中的形态特征对其进行甄选。首先从CCD图像中提取宇宙线像素，文中采用拉普拉斯边缘检测法从CCD原始图像中提取宇宙线候选像素列表。在提取后的宇宙线候选像素列表中包含一些坏像素和噪点，需要对这些坏像素和噪点进行处理。数据处理过程中，为了得到独立完整的宇宙线事件，采用凝聚层次聚类算法将宇宙线像素聚类成宇宙线事件。提取宇宙线事件的特征，结合μ子在CCD图像中的形态特征甄选出宇宙线μ子。在文中最后对甄选技术进行评估。

1 拉普拉斯边缘检测法提取宇宙线候选像素列表

在LAMOST数据中，单次曝光的CCD图像中含有上百条流量值很强的光纤光谱。这些CCD图像中，不仅包含所要观测的目标信息，同时还包括一些噪声（主要是宇宙线和坏像素）。在对CCD图像中的μ子进行甄选时，首先要从CCD图像中获取宇宙线候选像素列表。检测CCD图像宇宙线的方法有很多，例如经典的中值滤波方法[4]，拉普拉斯边缘检测法[5]，基于直方图的快速算法[6]以及万能噪声消除算法[7]等方法。本文采用改进的拉普拉斯边缘检测法[8]对CCD图像中的宇宙线候选像素列表进行提取。相较于其他方法，该方法在图像背景流量值很强的情况下，可以有效地从单次曝光的多光纤光谱中检测宇宙线。它是在参考文献[5]介绍的算法上进行改进,使得当宇宙线落在光纤光谱轮廓内时,宇宙线可以被识别出来。

改进的拉普拉斯边缘检测法的算法流程：

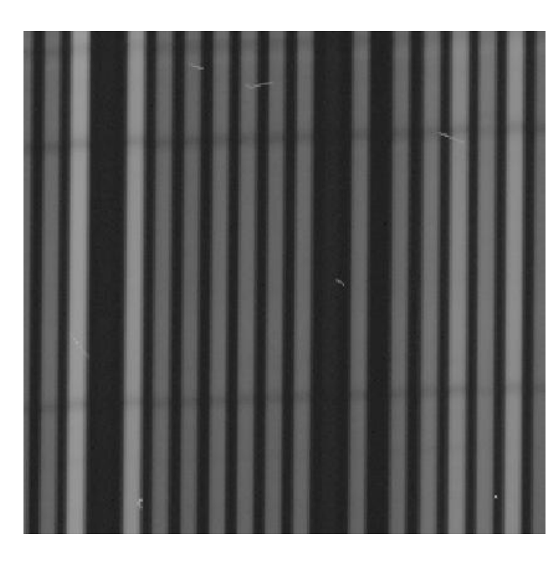
（1）对原图像进行预处理，然后与尺寸大小为3×3的Laplacian模板进行卷积运算得到图像I；结合泊松噪声和读出噪声，在原图像基础上构造剔除宇宙线的噪声模型N；再根据I和N的比率关系，生成一个原始的宇宙线候选像素列表；

（2）在原图像的基础上去除步骤（1）产生的宇宙线候选像素列表，生成一个新的图像。对图像中的每一条光纤进行处理，将距离光纤轨迹中心8个像素以内的像素点组成一个模

块。结合光谱在空间方向和色散方向的分布情况，拟合这些模块的图像轮廓（这些拟合的图像轮廓不包括宇宙线），得到拟合后的图像I_1；

（3）根据泊松噪声和读出噪声，对拟合后的图像I_1构造剔除宇宙线的噪声模型N_1；最后将原图像和拟合后图像的残差与噪声模型N_1进行比较生成宇宙线候选像素列表。

采用拉普拉斯边缘检测法提取宇宙线候选像素列表的结果见图 1。图 1(a)是原始图像，是从CCD图像中截取的一部分图像；图 1(b)对应于图 1(a)采用拉普拉斯边缘检测法提取后的宇宙线像素，图中的背景为黑色，对应的流量值为 0。从图 1 可以直观地显示出该算法可以有效地从CCD图像中提取宇宙线像素。



(a) 原始图像(290×278)



(b) 拉普拉斯边缘检测法提取的宇宙线图像(290×278)

图 1 拉普拉斯边缘检测法提取宇宙线

Fig.1 Extraction of cosmic rays with Laplacian edge detection algorithm

2 坏像素和噪点的处理

在采用拉普拉斯边缘检测法提取的宇宙线候选像素列表中，包括了宇宙线，CCD上的坏像素，以及算法导致的噪点。CCD上的坏像素是CCD本身的坏点，基本会在同一批拍摄的CCD图像中的同一位置处出现。可以利用同一目标的多次曝光图像来标识CCD上的坏像素，然后将它们去除。在使用拉普拉斯边缘检测法提取宇宙线候选像素列表时，如果图像中的光纤轨迹中心发生偏移时，会在拟合图像轮廓时产生较大的误差，最终在生成宇宙线列表的基础上产生拟合的偏差。这些偏差就是算法导致的噪点。这些噪点在图像中呈现几列虚线，虚线处的像素点个数远超过宇宙线在图像中每一列可能的像素点个数（虚线处的像素点个数范围为 150～500，而图像中每一列的宇宙线像素点个数范围为 8～50）。利用这一特点，在CCD图像中找出异常的列来标注算法导致的噪点，然后将它们从图像中去除。

3 凝聚层次聚类算法聚类宇宙线事件

从CCD图像中提取宇宙线像素后，采用凝聚层次聚类算法将宇宙线像素聚类成一个个独立的宇宙线事件。本文采用的算法是在经典的凝聚层次聚类算法上的改进。

层次聚类试图在不同层次对数据集进行划分，从而形成树形的聚类结构。数据集的划分可采用“自底向上”的聚合策略，也可采用“自顶向下”的分拆策略[9]。凝聚层次聚类（AGENS）是层次聚类中的一种，它对数据集的划分采用的是“自底向上”的聚合策略。它的主要思路是将数据集中的每一个样本看成单个的聚类簇，然后计算每个聚类簇之间的距离，找出其中距离最近的两个聚类簇将它们合并，重复上述距离计算和合并的过程，直至达到预设的聚类簇个数。

在经典的凝聚层次聚类算法中，算法的关键是计算类间距的方法和确定聚类簇的个数。每一个聚类簇是一个样本集合，可以采用计算集合的某种距离来计算类间距，常见的距离有最小距离，最大距离和平均距离。将宇宙线像素聚类成宇宙线事件，使得宇宙线事件之间相互独立，采用最小距离来计算类间距。最小距离计算公式见式（1），其中的C_i, C_j表示给定的聚类簇；$\{m_a\}_{a=1,2,...,k}$表示C_i的样本点集；$\{m_b\}_{b=1,2,...,l}$表示C_j的样本点集；dist是用于计算m_a, m_b样本点的距离函数。

$$d_{\min}(C_i, C_j) = \min_{m_a \in C_i, m_b \in C_j} dist(m_a, m_b) \tag{1}$$

在处理CCD图像时，无法预知图像中宇宙线事件的个数，所以无法预设聚类簇的个数。需要设定类间距d_{lim}作为终止运算条件。具体算法如下：

（1）假设数据集为D，数据集中的数据是CCD图像中宇宙线所对应的像素，将每个像素看作一个初始的聚类簇分别对应于$C_1, C_2, ..., C_n$。

$$D = \{m_1, m_2, ..., m_n\} \tag{2}$$

（2）计算所有聚类簇之间的类间距，式（3）采用欧氏距离和最小距离来计算类间距。如果两个聚类簇各自只含有一个样本点时，类间距是这两个样本点的欧氏距离；反之，类间距是两个聚类簇之间的最小距离。如果两个聚类簇的类间距不大于设定的阈值d_{lim}，将这两个聚类簇合并成一个聚类簇。

$$\mathrm{d} = \min_{m_a \in C_i, m_b \in C_j} \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \tag{3}$$

实验处理的数据是二维图像，式（3）中C_i, C_j表示数据集中的任意两个聚类簇，$\{m_a\}_{a=1,2,...,k}$表示C_i的样本点集；$\{m_b\}_{b=1,2,...,l}$表示C_j的样本点集，x_a和y_a对应于m_a在图像上的位置信息，x_b和y_b对应于m_b在图像上的位置信息。

（3）更新数据集；

（4）重复步骤（2）、（3），直到数据集中所有聚类簇的类间距大于设定的阈值，结束运算。

当CCD图像中出现大量宇宙线事件时，需要逐个计算聚类簇之间的距离，使计算量剧增，运行速率减慢。结合宇宙线事件是有计数的连续像素这一特点，通过判断与聚类簇最外层元素距离为d_{1im}的范围内是否存在其他的聚类簇，来优化算法。如果存在其他聚类簇，将其他聚类簇与该聚类簇合并；如果不存在，将数据集中未处理的聚类簇重复上述判断过程，直到所有聚类簇的类间距大于d_{1im}，结束运算。使用该方法将宇宙线像素聚类成宇宙线事件时，如果两个宇宙线事件的最小距离大于d_{1im}，说明两个宇宙线事件相互独立；如果最小距离不大于d_{1im}，说明两个宇宙事件是同一个宇宙线事件。在提取宇宙线像素的过程中，当宇宙线流量值和背景值比较接近时，有时会出现提取的宇宙线像素在图像中不连续。尽管这种情况很少出现，为减小宇宙线像素的缺失带来的误差，结合宇宙线在CCD图像中呈现随机稀疏分布的特性，设定d_{1im}为 3 对提取后的宇宙线像素聚类，可以有效地聚类宇宙线事件。

4 特征提取

将CCD图像中的宇宙线像素聚类成宇宙线事件后，结合μ子在图像中的形态特征，对宇宙线事件进行特征提取。在特征提取的过程中，主要内容是主成分分析和Pearson相关系数。通过主成分分析获取宇宙线事件的特征值，以及通过Pearson相关系数判断宇宙线事件是否呈线性，来甄选宇宙线事件中的μ子。

4.1 主成分分析（PCA）

主成分分析（PCA），也称为Karhunen-Loeve扩展，是一种经典的特征提取和数据表示技术[10]。参考文献[11]对主成分分析进行了总的概括。主成分分析（PCA）也可以用于对数据进行降维处理，目的是要找一组最优的标准正交向量基，使得数据投影到这组最优标准正交向量基上的方差最大。宇宙线事件是二维数据，假设在一个宇宙线事件中含有n个像素，其样本集$M=\{m_1, m_2, . . . , m_n\}$；将所有样本进行中心化，如式（4）。

$$\mathrm{m}_i^* = m_i - \frac{1}{\mathrm{n}}\sum_{i=1}^{n} m_i \tag{4}$$

定义样本集的散度矩阵为S_t：

$$S_t = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{m}_i^* - \overline{m})(m_i^* - \overline{m})^{\mathrm{T}} \tag{5}$$

其中样本集均值：

$$\overline{\mathrm{m}} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{m}_i^* \tag{6}$$

假设从高维空间投影到低维空间的投影矩阵为w，使投影后样本点方差最大化的优化目

标为：

$$S_t w = \lambda w \tag{7}$$

对S_t做特征值分解得到特征值λ，将特征值λ进行降序排序，选取最大的特征值。最大的特征值所对应的特征向量就是主成分的解，从而找到最优的标准正交向量基，得到宇宙线事件的特征向量。将宇宙线事件在特征向量方向上的投影作为宇宙线事件的特征。两个特征向量对应于两个特征值。其中一个特征值对应于宇宙线的长轴，即在CCD上宇宙线直线径迹所对应的轴；另一个特征值对应于宇宙线的短轴，即宇宙线在CCD图像上的宽度。μ子在CCD图像的径迹是直线型，且径迹长度无法确定，但宽度有限，是个小量。选取标记宽度的特征作为μ子的特征，并将其记为“投影宽度”。

4.2 Pearson相关系数

相关系数是用来衡量变量之间线性相关程度的统计量。用于定义相关系数的方式有很多，最为常用的是Pearson相关系数。Pearson相关系数用于描述两个变量之间是否存在线性关系。参考文献[12]对Pearson相关系数进行了详细介绍。假设有两个变量分别是x和y, 它们的Pearson相关系数用r表示，则有：

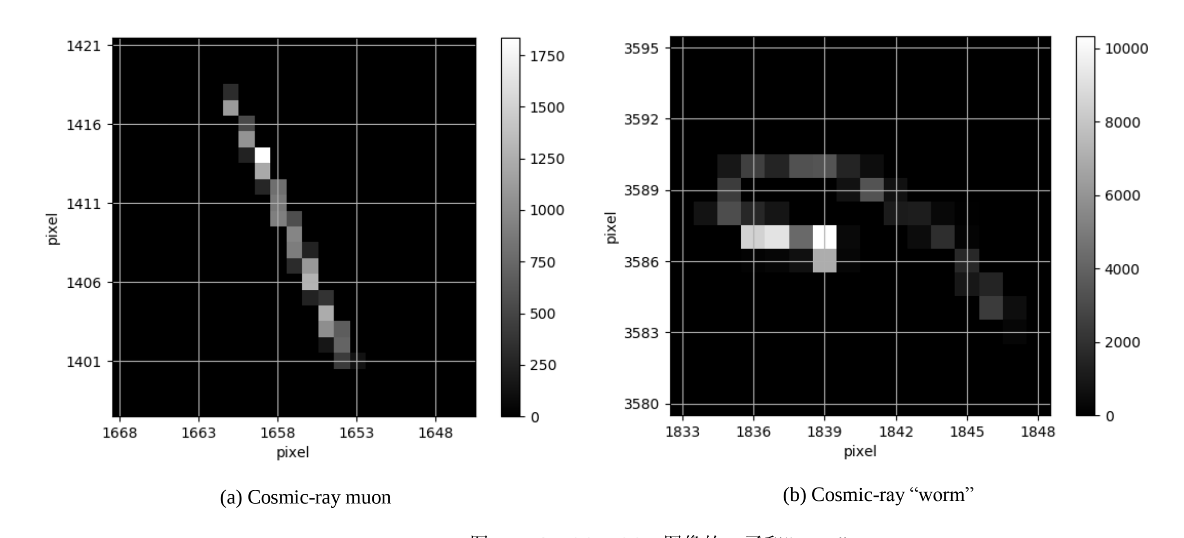
[illegible] (8)

其中的μ_x，μ_y分别对应于变量x，y的均值。r的取值范围为[-1, 1]，$|r|$越接近于1时，两个变量之间的线性相关性质越显著。当$r>0$时，两个变量之间呈正相关；当$r<0$时，两个变量之间呈负相关；当$r=0$时，两个变量之间彼此独立。

宇宙线μ子在CCD上呈直线轨迹，则μ子在CCD图像中的横纵坐标存在线性关系。利用Pearson相关系数来标记μ子的这一特征。由于μ子在图像中的位置有时会出现正相关，有时会出现负相关，为了显示线性相关性质的显著程度，在传统的Pearson相关系数的基础上取绝对值。下文提到的Pearson相关系数r对应取绝对值的Pearson相关系数。

5 实验结果及分析

利用宇宙线μ子与其他粒子在CCD图像中的形态特征差异来甄选μ子。在CCD图像中最为常见的粒子是μ子和“worms”（“worms”是多级散射的电子,详细介绍见参考文献[2]）。α粒子在CCD图像上极少出现。CCD图像中还存在一些形状奇特的宇宙线事件，因无法确定其形态特征，难以判断属于何种宇宙线粒子，在下文中未做讨论。图 2 是从LAMOST数据中获得的图像，图2(a)是宇宙线中的μ子，图2(b)是宇宙线中的“worm”。如图2(b)所示，“worm”形状呈现一定的弧度。



(a) Cosmic-ray muon　　　　(b) Cosmic-ray “worm”

图 2　LAMOST CCD 图像的 μ 子和“worm”

Fig.2 cosmic-rays muon and “worm” in the CCD image of LAMOST

利用文中章节 1 至章节 3 的处理过程，我们从 2015 年的LAMOST CCD图像中随机选取了 200 多张CCD图像，得到 3 万多个宇宙线事件，将选取的宇宙线事件的位置索引信息作为样本集*C1*。宇宙线事件的位置索引信息可以反映出宇宙线事件在CCD图像上呈现的形态。从样本集中通过人眼识别的方式，选取 1000 个μ子和 1000 个“worms”作为数据样本 1。统计这些宇宙线事件的“投影宽度”和Pearson相关系数，结果见表 1。表 1 中宇宙线事件“投影宽度”的均值记为μ_{wp}，标准差记为σ_{wp}；Pearson相关系数的均值记为μ_r，标准差记为σ_r。

表 1 关于“投影宽度”和Pearson相关系数，宇宙线μ子、“worms”的特征对比

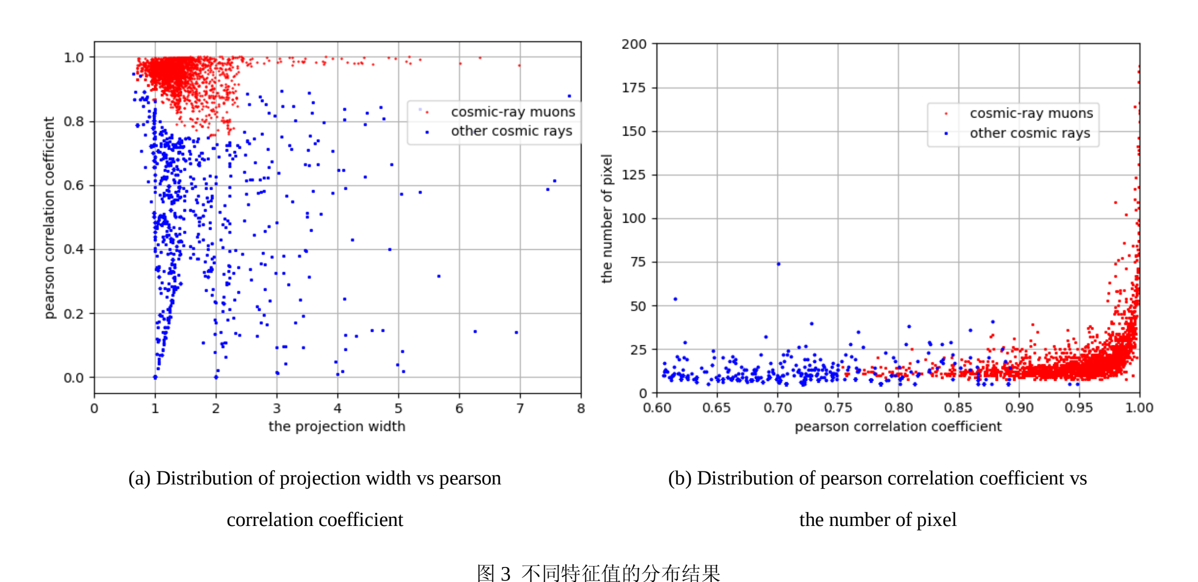
Tab.1 Feature comparison between muons and “worms” in cosmic rays of the “Projection width” and the Pearson correlation coefficient

Type of cosmic rays	“The projection width”($\mu_{wp}\pm\sigma_{wp}$)	Pearson correlation coefficient($\mu_r\pm\sigma_r$)
muons	1.43±0.50	0.93±0.06
“worms”	5.10±2.03	0.73±0.15

根据表 1 可以直观的看到μ子和“worms”的区别。在“投影宽度”中，μ子的特征值小，而“worms”的特征值大；在Pearson相关系数中，由于μ子在CCD中呈直线径迹，对应的横纵坐标的线性相关程度更显著，对应的Pearson相关系数值更接近 1，而“worms”由于在CCD中发生多次散射，径迹有一定的弧度，对应的Pearson相关系数值较μ子的更小。

从样本集*C1* 中，通过人眼识别的方式，随机选取 2500 个μ子和 2500 个不是μ子的其他宇宙线事件（包含“worms”、X射线、可能的α粒子等其他粒子）作为数据样本 2。这 5000 个宇宙线事件的“投影宽度”和Pearson相关系数的分布结果如图 3(a)所示。在图 3(a)中，利

用“投影宽度”和Pearson相关系数两个特征可以将绝大多数的μ子和其他宇宙线区分开。在图中Pearson相关系数大于 0.8 和“投影宽度”小于 1 的范围内，μ子和其他宇宙线事件不能很好的区分；观测这部分与μ子特征值相近的宇宙线事件，发现该部分宇宙线事件的像素个数很少。分析Pearson相关系数和宇宙线事件像素个数的关联，二者的分布结果见图 3(b)。图 3(a)中不易区分的μ子和其他宇宙线事件，在图 3(b)中可以通过宇宙线事件的像素个数有效地区分。



(a) Distribution of projection width vs pearson correlation coefficient

(b) Distribution of pearson correlation coefficient vs the number of pixel

图 3 不同特征值的分布结果

Fig.3 The distribution of different characteristic values

对于像素个数少的宇宙线事件，只结合“投影宽度”和Pearson相关系数甄选μ子会有一定的误差。当甄选μ子时，还需要结合宇宙线事件的像素个数。图 4 显示了数据样本 2 中宇宙线事件像素个数的统计结果。横坐标表示宇宙线事件的像素点个数，纵坐标表示像素点个数对应的宇宙线事件的个数。图 4(a)是整个数据集的统计结果。图中宇宙线事件的像素点个数越多，与之对应的宇宙线事件个数越少。基于图 4(a)，选取像素点个数不大于 30 的宇宙线事件进行统计，结果见图 4(b)。图中，宇宙线μ子的像素点个数大于 8，其他宇宙线事件的像素点个数大于 5。在非μ子的其他宇宙线中，将近一半的宇宙线的像素点个数小于 8。在像素点个数大于 8 的情况下，μ子的个数比其他宇宙线的个数多。因此将像素点个数设置为 8，对宇宙线中μ子和其他宇宙线事件进行分类。

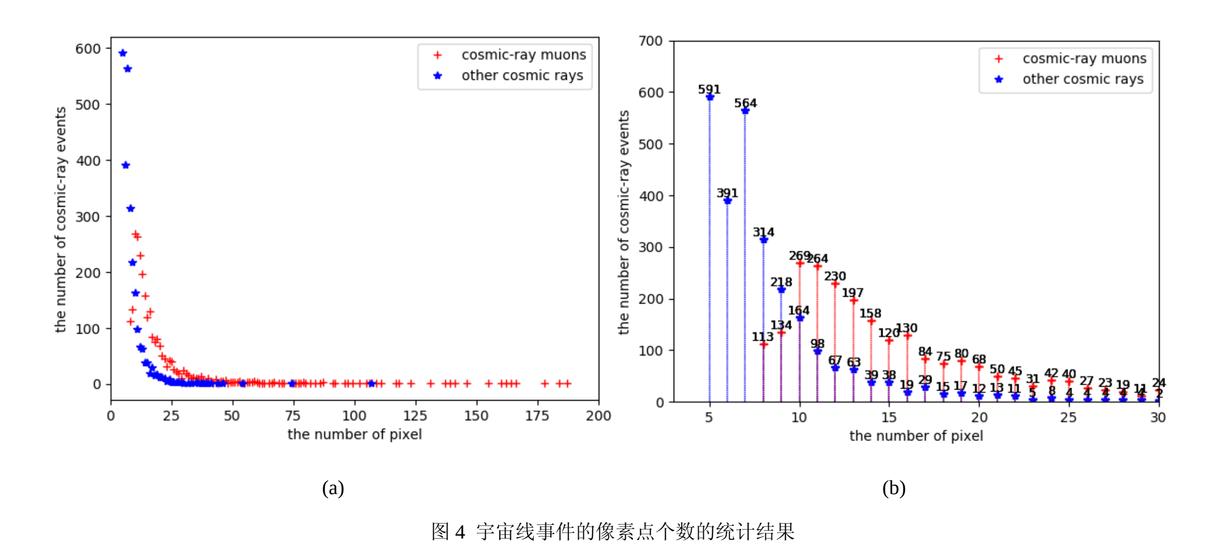


图 4 宇宙线事件的像素点个数的统计结果

Fig.4 The statistical result of the number of pixels in cosmic-ray event

根据宇宙线事件特征值的不同分布，设定不同的特征值参数，对 μ 子进行甄选。从样本集*C1* 中，随机选取 2500 个 μ 子和 2500 个不是 μ 子的宇宙线在CCD上的位置索引信息作为测试数据集C2，设定不同的特征值参数对宇宙线中的μ子进行甄选，最后的甄选结果如表 2 所示。表 2 显示了甄选结果的准确性，表中的第一列是特征参数的设定，其中的μ_{mw}对应于μ子“投影宽度”的均值，σ_{mw}对应于μ子“投影宽度”的标准差，μ_{mr}对应于μ子Pearson相关系数的均值，σ_{mr}对应于μ子Pearson相关系数的标准差；第二列是正确甄选宇宙线中μ子的个数；第三列是被错分成μ子的宇宙线事件个数；第四列是甄选结果的正确性；最后一列是甄选结果的灵敏度，用于衡量算法对μ子的识别能力。假设在整个数据集中，μ子的个数是*M*，非μ子的宇宙线事件个数为*F*。在甄选结果中，被正确甄选为μ子的宇宙线事件个数为*TM*, 被正确甄选为非μ子的宇宙线事件个数为*TF*，则甄选结果的准确率公式见式(9)，甄选结果的灵敏度公式见式(10)。

$$Accur = \frac{TM}{MF} \quad (9)$$

[illegible] (10)

从表 2 可以看出：当对“投影宽度”、Pearson相关系数和宇宙线事件的像素个数设定不同阈值时，会得到不同的准确率；当特征值设定合适的阈值时，可以有效地从宇宙线事件中甄选μ子。表 2 特征参数的选择，需要结合表 1 中μ子不同特征的均值和标准差。通过统计宇宙线μ子的“投影宽度”和Pearson相关系数，得到描述不同特征的具体数值参数以及参数的波动情况。结合不同特征的参数波动，对测试数据集中的μ子进行甄选，得到最终的甄选结果。

表 2 设定不同特征参数时，μ子的甄选结果

Tab.2 The results of cosmic-ray muons selection when setting different feature parameters

Condition about "Projection width"(pw) , Pearson correlation coefficient(r) and the number of cosmic event(pixel)	Numbers of muons detected correctly	Numbers of muons detected incorrectly	Accuracy	Sensitive
(($r \geq \mu_{mr}-3\sigma_{mr}$ and $0<$ pw$<\mu_{mw}+\sigma_{mw}$) or ($r \geq \mu_{mr}$ and pw$\neq$0)) and pixel>=8	2412	88	0.965	0.965
(($r \geq \mu_{mr}-3\sigma_{mr}$ and $0<$ pw$<\mu_{mw}+2\sigma_{mw}$) or ($r \geq \mu_{mr}$ and pw$\neq$0)) and pixel>=8	2500	108	0.978	1
(($r \geq \mu_{mr}-3\sigma_{mr}$ and $0<$ pw$<\mu_{mw}+3\sigma_{mw}$) or ($r \geq \mu_{mr}$ and pw$\neq$0)) and pixel>=8	2500	117	0.977	1

当对宇宙线特征设置合适阈值时，得到的准确率约为 0.98，说明该方法可以有效地甄选宇宙线中的μ子。该方法对μ子的识别度很高，但是会将与μ子特征相似、不是μ子的宇宙线事件错分成μ子。

尽管该算法在数据处理过程中会存在误差，但是通过分析μ子的本征特征，结合数据集中μ子和其他宇宙线事件的形态差异，依然可以有效地甄选宇宙线中的μ子。该方法的核心是特征参数的设定。如果特征值的参数设定恰当，最后的甄选结果会更可观，也可以通过训练大量数据集的方式来调整特征参数，优化方法。根据表 2 中最优的甄选条件（表 2 中第二行所对应的特征参数），对LAMOST 6 号红端的CCD相机，在 2017 年 11 月份拍摄的图像进行处理。统计这一个月，曝光时间为 600.0s的CCD图像中μ子的个数。统计结果见图 5。图中横坐标表示拍摄CCD图像的mjm（修正儒略分钟数），纵坐标表示CCD图像中μ子的个数。图中蓝色竖线对应于第二横坐标拍摄CCD图像的日期。图中直观地显示出宇宙线μ子在这一个月的涨落情况。

6 总结

本文探讨了一种简单有效地对单次曝光CCD图像中的μ子进行甄选的方法，结合宇宙线在图像上的形态特征，达到高效处理数据的效果。该方法使用拉普拉斯边缘检测法提取宇宙线像素；采用聚类算法将图像中的宇宙线像素聚类成宇宙线事件；聚类得到宇宙线事件后，通过特征提取和宇宙线事件之间的形态特征差异对μ子进行甄选。在这个过程中，甄选准确率依赖于特征参数的设定，可以通过不断的统计和测试，设定更优的参数以及增加提取的特征，来提高甄选结果的准确性。

在获取训练和测试样本集的过程中，通过人眼识别的方式来区分宇宙线μ子和其他宇宙线事件，会导致选择的数据样本集有一定的偏差，影响之后的统计和甄选结果。

致谢：感谢中国科学院高能物理研究所的胡红波老师的有益讨论。

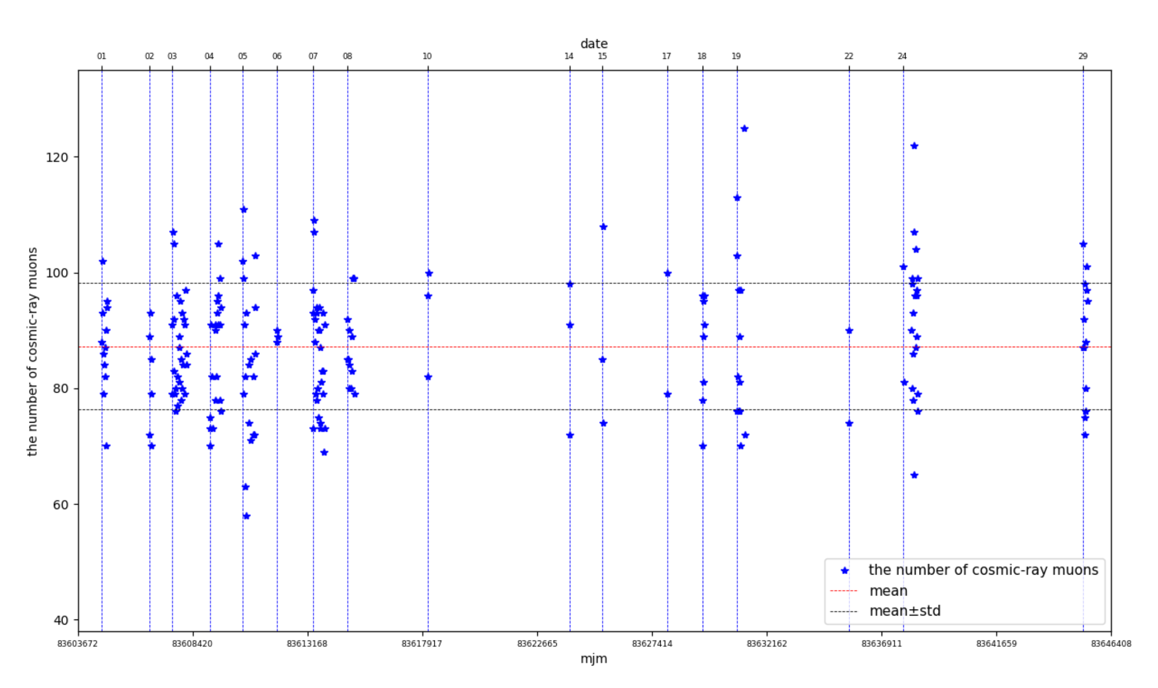


图 5 2017 年 11 月份 LAMOST 6r CCD 图像中宇宙线 μ 子个数的统计结果

Fig.5 The statistical result of the numbers of cosmic-ray muons in 6r CCD of LAMOST in November 2017

参考文献:

[1] O.C.阿尔科费尔.宇宙线入门[M].徐春娴，朱清棋，译.北京：科学出版社,1987:34.

O.C.Allkofer. Introduction to Cosmic Radiation[M]. C.X. Xu & Q.Q. Zhu,Trans. Beijing: Science Press, 1987:34.

[2] Don Groom. Cosmic Rays and Other Nonsense in Astronomical CCD Imagers[J]. Experimental Astronomy, 2002, 14(1): 45 -55.

[3] Cui,X.Q., Zhao,Y.H., Chu,Y.Q., etal. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope(LAMOST)[J]. RAA, 2012, 12(9): 1197 -1242.

[4] Gonzalez RC, Woods RE. 数字图像处理[M]. 阮秋琦，阮宇智等译.北京：电子工业出版社，2007:185-186.

Gonzalez RC, Woods RE. Digital Image Processing[M]. Q.Q. Ruan & Y.Z. Ruan, Trans. Beijing: Electronic Industry Press, 2007: 185-186.

[5] Pieter G. van Dokkum. Cosmic-Ray Rejection by Laplacian Edge Detection[J]. PASP, 2001, 113: 1420 -1427.

[6] Pych W. A Fast Algorithm for Cosmic-ray Removal from Single Images[J]. PASP, 2004, 116(816): 148 -153.

[7] Garnett R,Huegerich T, Chui C, etal. A Universal Noise Removal Algorithm with An Impulse Detector[J]. IEEE Transaction On Image Processing. 2005, 14(11): 1747 -1754.

[8] Zhongrui Bai, Haotong Zhang, etal. Cosmic Ray Removal in Fiber Spectroscopic Image[J]. PASP, 2017, 129(972): 024004(11pp)

[9] 周志华. 机器学习[M]. 北京：清华大学出版社，2016: 214.

Zhou Zhihua. Machine Learning[M]. Beijing: tsinghua university press, 2016: 214.

[10] C.Li, Y. Diao, H. Ma and Y. Li, A Statistical PCA Method for Face Recognition[J].Intelligent Information Technology Application, 2008, pp. 376-380.

[11] Sasan Karamizadeh, Shahidan M. Abdullah. An Overview of Principal Component Analysis[J]. Journal of Signal and Information Processing, 2013, 4: 173-175.

[12] Leo Egghe, Loet Leydesdorff. The relation between Pearson's correlation coefficient r and Salton's cosine measure[J]. Journal of the American Society for Information Science & Technology, 2009, 5: 1027- 1036.

Cosmic-ray Muons Extraction Technology in CCD Image

Feng Haixia[1,2,3], Chen Jianjun[1,3*], Deng Jianrong[1], Zhao Yongheng[1,2,3]

(1.National Astronomical Observatories，Chinese Academy of Science, Beijing 100101, China,Email: hxfeng@nao.cas.cn;

2. Chinese Academy of Science，Beijing 100049, China;

3. Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Science, Beijing 100101, China,

Email: jjchen@nao.cas.cn)

Abstract: The cosmic rays in CCD(Charge Coupled Device) image are formed by high-energy particles from outer space which pass through the atmosphere, collide with particles in the atmosphere and fall on the CCD panel finally. The cosmic-ray muons are the main component of cosmic rays in the atmosphere. In order to study cosmic-ray muons' property and change rule, we need extract cosmic-ray muons in CCD images. This paper introduces an easy and effective method to extract cosmic-ray muons in CCD images. The main work includes the following steps: searching the cosmic rays candidate lists from CCD images by the Laplacian edge detection algorithm; removing bad pixels and noise from the extracted cosmic rays candidate lists; clustering cosmic rays pixels into cosmic ray events by AGglomerative NESting method; then extracting cosmic-ray muons according to the extracted features. Finally, the result of extracting cosmic-ray muons is summarized and analyzed.

Key words: CCD; Cosmic-ray muons; Laplacian edge detection algorithm; AGglomerative NESting; Feature extracting

*Corresponding author